

MANUEL TECHNIQUE

&

GUIDE UTILISATEUR

TEST DE RAISONNEMENT 20'

Ce manuel a été réalisé par les psychologues du département de Recherche et Innovation de Central Test.

Le TEST DE RAISONNEMENT '20 et le logo Central Test sont des marques déposées.

Tout droit de reproduction, modification ou réutilisation doit être accordé par écrit par l'éditeur Central Test.

Copyright © 2020 par Central Test - Empowering Talent.

Sommaire

Introduction	3
1. Description générale et domaines d'application	4
1.1. Description des facteurs	4
1.2. Principaux domaines d'application	5
1.3. Structure du questionnaire et mode d'administration	5
1.4. Questions techniques et de sécurité	6
2. Propriétés psychométriques	7
2.1. Descriptif de l'échantillon	7
2.2. Normes	8
2.3. Fidélité et Validité	12
2.4. TEST DE RAISONNEMENT 20' et égalité des chances	14
3. Informations supplémentaires et Considérations	18
Références	19

Introduction

Wechsler décrit l'intelligence comme “la capacité globale de l'individu d'agir délibérément, de penser de manière rationnelle et de gérer efficacement son environnement”.

Cette capacité est indispensable en entreprise car c'est elle qui nous permet de comprendre et d'intégrer les consignes, de réfléchir, d'apprendre. Afin d'aider les entreprises à évaluer ces capacités chez leurs candidats nous avons développé le **TEST DE RAISONNEMENT 20'**. Ce test se focalise sur trois domaines de l'intelligence : le raisonnement figural, le raisonnement verbal et le raisonnement numérique. On considère que ces trois domaines sont importants car ils englobent des compétences pertinentes dans le monde de l'entreprise, telles que la communication, la planification, l'organisation, le calcul et la prévision des résultats.

1. Description générale et domaines d'application

L'intelligence est un concept à plusieurs facettes et multicouches. Parmi tous les facteurs considérés comme constituant l'intelligence, les chercheurs en ont découvert trois principaux, à savoir les capacités de raisonnement numérique, verbal et figural, qui permettent de mieux prédire l'intelligence générale ou le "facteur g" de Spearman. Le quotient intellectuel (QI) obtenu lors de la passation de certains tests d'intelligence est une représentation numérique de ce "facteur g". Il offre une image globale de la capacité d'un individu à raisonner, planifier, résoudre des problèmes, penser de manière abstraite, comprendre des idées complexes, apprendre rapidement, tirer des leçons de son expérience et s'adapter à la plupart des situations.

Par conséquent, le **TEST DE RAISONNEMENT 20'** est conçu pour évaluer ces trois facteurs essentiels de l'intelligence - *Raisonnement figural*, *Raisonnement verbal* et *Raisonnement numérique*.

Le rapport du **TEST DE RAISONNEMENT 20'** comprend les éléments suivants :

- Un quotient intellectuel, ramené sur une échelle de 0 à 20, reflétant les performances globales d'un individu
- Le nombre de réponses correctes obtenues par l'individu
- Une représentation graphique des scores pour chacun des trois facteurs - *Raisonnement figural*, *Raisonnement Verbal* et *Raisonnement Numérique*
- Une description détaillée de chaque facteur
- La présentation des solutions si le client le souhaite, où il a accès aux réponses individuelles des réponses correctes.

1.1. Description des facteurs

Le **TEST DE RAISONNEMENT 20'** mesure trois des composantes les plus importantes de l'intelligence, à savoir *Raisonnement figural*, *Raisonnement verbal* et *Raisonnement numérique*.

Raisonnement figural - Le facteur de *Raisonnement figural* fait référence à la faculté à établir des liens entre divers éléments et à découvrir les relations existantes au sein d'ensembles plus ou moins complexes.

Ce type de raisonnement intervient dans quasiment toutes les situations de la vie courante. Chaque fois qu'une personne doit découvrir les relations unissant diverses parties d'un ensemble, les facultés logiques sont sollicitées. C'est la capacité la plus fortement liée à l'intelligence générale.

Dans la vie professionnelle, de nombreuses activités font appel au raisonnement de type figural. À titre d'exemple, l'informatique, le droit, la médecine sont des activités dans lesquelles la logique est prédominante.

Raisonnement verbal - Le facteur de *Raisonnement verbal* représente la faculté à comprendre les subtilités de la langue et à manier les données verbales avec justesse.

Les situations faisant appel au facteur Verbal sont nombreuses et des plus variées : la maîtrise du discours, l'argumentation, l'humour, etc.

Dans une culture où l'écrit et le langage occupent une place considérable, l'aptitude à identifier et à comprendre rapidement les informations qui nous environnent constitue un véritable atout dans la vie quotidienne.

Raisonnement numérique – Le facteur de *Raisonnement numérique* représente la faculté à raisonner sur les données numériques avec justesse.

Les situations faisant appel au facteur Numérique sont nombreuses et d'une grande variété : calcul mental, économie, finances, comptabilité, négociations commerciales, interprétation de données chiffrées, etc.

1.2. Principaux domaines d'application

Le **TEST DE RAISONNEMENT 20'** est recommandé pour les professionnels tels que les commerciaux, les entrepreneurs, les directeurs et autres employés de bureau.

Le **TEST DE RAISONNEMENT 20'** est spécialement réservé aux professionnels des départements des ressources humaines des organisations, écoles et collèges, ainsi qu'aux formateurs.

1.3. Structure du questionnaire et mode d'administration

Format du TEST DE RAISONNEMENT 20'

- Nombre de questions : 21
- Format de l'évaluation : Questions à choix multiples
- Durée : 20 minutes (Chronométré)
- Administration : En ligne

Administration du TEST DE RAISONNEMENT 20'

Il est recommandé que le **TEST DE RAISONNEMENT 20'** soit administré sous surveillance. Toutefois, dans le cas d'une administration non supervisée, l'administrateur du test doit prendre les précautions nécessaires pour minimiser la possibilité, pour le candidat, d'obtenir une aide externe.

1.4. Questions techniques et de sécurité

Les points suivants servent de directives pour administrer en toute sécurité le **TEST DE RAISONNEMENT 20'**. Celles-ci sont utiles pour garantir que le **TEST DE RAISONNEMENT 20'** est administré de manière objective et équitable. Les spécifications de sécurité sont également importantes pour la protection des informations du candidat, à la fois pendant l'administration et après l'évaluation.

Respecter la Charte Éthique

En tant qu'éditeur d'outils d'évaluation en ligne, *Central Test* respecte les conventions légales du droit du travail français et les dispositions réglementaires issues de la loi du 6 janvier 1978 relative à l'informatique, aux fichiers et aux libertés : Déclaration N°1834067. De plus, depuis le 25 mai 2018, *Central Test* est conforme aux réglementations de la RGPD (Règlement général sur la protection des données) édictées par le Parlement et le Conseil européens.

Les résultats du test et les données personnelles ne peuvent être transmis à des tiers sans le consentement explicite du candidat. Les candidats ont le droit d'accéder à leurs résultats et l'évaluateur doit leur accorder cet accès si demandé. En outre, si un répondant l'exige, toutes les données devront être définitivement supprimées. Il devra pour ce faire contacter l'entreprise ou *Central Test* directement.

Confidentialité et Sécurité

Central Test garantit la sécurité et la confidentialité des données générées durant l'administration des tests en ligne. En effet, les codes d'accès sont cryptés dans la base de données et ne sont accessibles à personne. Les mots de passe sont générés par ordinateur et ne sont pas visibles par les entreprises. Un module de réécriture d'URL (générateur de liens sécurisés) est installé et assure le passage transparent des paramètres pour l'utilisateur.

La navigation et l'échange de données via l'application *Central Test* peuvent être sécurisés et représentés par une icône en forme de cadenas dans le navigateur Internet. L'application de *Central Test* dispose d'un certificat <<GeoTrust SSL TRUE BusinessID>>, une signature numérique émise et certifiée par GeoTrust®. Ce certificat prouve que nous prenons la sécurité au sérieux et il protège toutes les évaluations de *Central Test* contre le phishing.

Les invitations au test sont envoyées par des liens sécurisés d'utilisation limitée (en nombre et en durée). La durée du test et les adresses IP sont enregistrées.

2. Propriétés psychométriques

Central Test s'engage à veiller à ce que les données fournies par ses évaluations soient objectives et à ne pas discriminer les personnes en fonction de facteurs qui ne devraient pas être pris en compte lors de l'embauche, notamment la race, le genre et l'âge. C'est également le cas du **TEST DE RAISONNEMENT 20'** et ce manuel prend spécifiquement en compte un échantillon francophone afin de garantir un groupe de référence approprié. Nos statisticiens et chercheurs se sont assurés des qualités métriques du test pour l'ensemble des populations auxquelles il est destiné. Nous présentons ici les données pour le marché francophone.

L'échantillon comprend 7 754 professionnels francophones originaires de 27 pays différents.

2.1. Descriptif de l'échantillon

Tableau 1. Échantillon par genre.

Genre	Pourcentage (%)	N
Femme	48,92	3793
Homme	51,08	3961

Tableau 2. Échantillon par groupe d'âge.

Groupe d'âge	Pourcentage (%)	N
Non précisé	0,013	1
18 – 25	32,87	2549
26 – 33	33,05	2563
34 – 41	19,63	1522
42 – 49	7,69	596
50+	4,16	323

Tableau 3. Descriptif de l'échantillon selon le pays d'origine.

Pays d'origine	Pourcentage (%)	N
Autres	7%	563
France	62%	4795
Tunisie	19%	1477
Congo	10%	753
Madagascar	2%	166

Note : Les pays dans la catégorie « Autres » ont été regroupés car ils représentent un nombre non suffisant pour être représentatif. Cela concerne le Canada, la Belgique, le Maroc, l'Algérie, l'Afrique du sud, l'Angola, le Cameroun, les Comores, la Côte d'Ivoire, le Gabon, le Ghana, la Guinée, l'île Maurice, le Mali, la Mauritanie, le Mozambique, le Niger, le Nigéria, la République Centre-Africaine, le Rwanda, le Sénégal, le Tchad et le Togo.

2.2. Normes

Les normes nous aident à comparer le score d'un individu à un test avec les scores d'autres personnes remplissant la même évaluation, généralement dans des conditions similaires. Plus la correspondance entre les caractéristiques de l'échantillon et celle de la population cible est étroite, plus la distribution sera précise comme guide de classement. Grâce aux normes, les concepteurs de tests, les psychologues et les utilisateurs sont en mesure d'établir des comparaisons entre individus et de prendre des décisions plus objectives. Les groupes de référence nous permettent d'attribuer une signification aux tests et de quantifier les performances d'un individu, le niveau d'un attribut ou trait par rapport à ses pairs.

Les tableaux 4 à 7 détaillent les moyennes, écarts-types (SD), médianes, minima (Min), maxima (Max), étendues et estimations standards (SE) de l'échantillon francophone sur les facteurs du **TEST DE RAISONNEMENT 20'**. Ces tableaux sont divisés en catégories en fonction du genre, du groupe d'âge et du pays d'origine.

Tableau 4. Statistiques descriptives pour l'ensemble de la population.

Facteur	Moyenne	SD	Médiane	Min	Max	Etendue	SE
Raisonnement figural	6,32	2,34	7,10	0,00	10,00	10,00	0,03
Raisonnement numérique	5,77	2,63	5,70	0,00	10,00	10,00	0,03
Raisonnement verbal	6,38	2,32	7,10	0,00	10,00	10,00	0,03

Tableau 5. Score par facteur en fonction du groupe d'âge.

Facteur	Moyenne	SD	Médiane	Min	Max	Etendue	SE
18 – 25							
Raisonnement figural	6,66	2,17	7,10	0,00	10,00	10,00	0,04
Raisonnement numérique	5,88	2,56	5,70	0,00	10,00	10,00	0,05
Raisonnement verbal	6,75	2,16	7,10	0,00	10,00	10,00	0,04

Facteur	Moyenne	SD	Médiane	Min	Max	Etendue	SE
26 – 33							
Raisonnement figural	6,28	2,43	5,70	0,00	10,00	10,00	0,05
Raisonnement numérique	5,76	2,72	5,70	0,00	10,00	10,00	0,05
Raisonnement verbal	6,10	2,41	5,70	0,00	10,00	10,00	0,05

Facteur	Moyenne	SD	Médiane	Min	Max	Etendue	SE
34 – 41							
Raisonnement figural	6,03	2,36	5,70	0,00	10,00	10,00	0,06
Raisonnement numérique	5,68	2,59	5,70	0,00	10,00	10,00	0,07

Raisonnement verbal	6,08	2,35	5,70	0,00	10,00	10,00	0,06
---------------------	------	------	------	------	-------	-------	------

Facteur	Moyenne	SD	Médiane	Min	Max	Etendue	SE
---------	---------	----	---------	-----	-----	---------	----

42 – 49

Raisonnement figural	6,16	2,44	5,70	0,00	10,00	10,00	0,10
Raisonnement numérique	5,86	2,72	5,70	0,00	10,00	10,00	0,11
Raisonnement verbal	6,63	2,28	7,10	0,00	10,00	10,00	0,10

Facteur	Moyenne	SD	Médiane	Min	Max	Etendue	SE
---------	---------	----	---------	-----	-----	---------	----

50+

Raisonnement figural	5,85	2,33	5,70	0,00	10,00	10,00	0,13
Raisonnement numérique	5,65	2,63	5,70	0,00	10,00	10,00	0,15
Raisonnement verbal	6,82	2,26	7,10	0,00	10,00	10,00	0,13

Tableau 6. Score par facteur en fonction du genre.

Facteur	Moyenne	SD	Médiane	Min	Max	Etendue	SE
---------	---------	----	---------	-----	-----	---------	----

Femmes

Raisonnement figural	6,22	2,31	5,70	0,00	10,00	10,00	0,04
Raisonnement numérique	5,55	2,57	5,70	0,00	10,00	10,00	0,04
Raisonnement verbal	6,35	2,32	7,10	0,00	10,00	10,00	0,04

Facteur	Moyenne	SD	Médiane	Min	Max	Etendue	SE
---------	---------	----	---------	-----	-----	---------	----

Hommes

Raisonnement figural	6,41	2,36	7,10	0,00	10,00	10,00	0,04
----------------------	------	------	------	------	-------	-------	------

Raisonnement numérique	5,97	2,68	5,70	0,00	10,00	10,00	0,04
Raisonnement verbal	6,41	2,32	7,10	0,00	10,00	10,00	0,04

Tableau 7. Score par facteur selon le pays d'origine.

Facteur	Moyenne	SD	Médiane	Min	Max	Etendue	SE
<i>Tunisie</i>							
Raisonnement figural	5,91	2,18	5,70	0,00	10,00	10,00	0,06
Raisonnement numérique	5,90	2,47	5,70	0,00	10,00	10,00	0,06
Raisonnement verbal	5,12	2,13	5,70	0,00	10,00	10,00	0,05

Facteur	Moyenne	SD	Médiane	Min	Max	Etendue	SE
<i>Madagascar</i>							
Raisonnement figural	6,32	2,06	5,70	1,40	10,00	8,60	0,16
Raisonnement numérique	5,94	2,61	5,70	0,00	10,00	10,00	0,20
Raisonnement verbal	5,75	2,05	5,70	0,00	10,00	10,00	0,16

Facteur	Moyenne	SD	Médiane	Min	Max	Etendue	SE
<i>Congo</i>							
Raisonnement figural	3,77	1,88	4,30	0,00	10,00	10,00	0,07
Raisonnement numérique	3,07	1,98	2,90	0,00	10,00	10,00	0,07
Raisonnement verbal	4,12	1,78	4,30	0,00	10,00	10,00	0,06

2.3. Fidélité et Validité

Fidélité

La fidélité indique dans quelle mesure une évaluation ne comprend pas d'erreur et le degré auquel les items d'une échelle mesurent le même construit. Par exemple, tous les items qui évaluent le raisonnement numérique sont cohérents et reflètent la définition de ce facteur. Une première préoccupation, lorsque l'on élabore et utilise les évaluations, est de savoir à quel point les erreurs de la mesure aléatoire influencent la performance du test. Si un test n'est pas fiable, il est impossible de le valider car l'exactitude des résultats (validité) dépend de la cohérence (fidélité) avec laquelle ces mêmes résultats sont produits. La fidélité permet aux utilisateurs du test d'être assurés que les résultats reflètent précisément les attitudes, aptitudes et comportements au travail des répondants. Lorsqu'une évaluation a une très bonne fidélité, elle contribue également à empêcher les biais susceptibles d'être provoqués par des erreurs de jugement durant le processus de recrutement et de sélection.

Les valeurs de fidélité varient de 0 à 1, les plus élevées étant plus souhaitables. En règle générale, une fidélité de 0,70 ou plus est recommandée (Nunnally, 1978), avec des valeurs supérieures à 0,9 (excellent) ; 0,7 à 0,9 (bon) ; 0,6 à 0,7 (acceptable) ; 0,5 à 0,6 (faible) et des valeurs inférieures à 0,5 (inacceptable).

La fidélité interne ou inter-item indique dans quelle mesure les items d'une échelle sont homogènes dans leur contenu et mesurent tous le même construit. Pour le **TEST DE RAISONNEMENT 20'**, un estimateur puissant et moins connu de la fidélité interne basé sur les hypothèses de la théorie du test classique, appelé la plus grande limite inférieure (GLB) a été utilisé. La GLB produit de meilleurs résultats sur l'alpha de Cronbach (Oliver et al., 2014). Cependant, dans les cas où les données ne sont pas distribuées normalement, il produit un meilleur résultat sur, ensemble, l'alpha et l'oméga (Wilcox et al., 2014), c'est à dire la fidélité composite.

Pour le **TEST DE RAISONNEMENT 20'**, la limite de fidélité la plus basse obtenue a été de 0,81, ce qui indique que l'échelle a un bon niveau de cohérence et est fiable.

Validité

La validité correspond à la mesure dans laquelle un test évalue le construit sous-jacent, dans le cas présent les capacités de raisonnement. En termes de validité psychométrique, cela se réfère au degré auquel les faits et la théorie scientifique étayent l'interprétation des scores aux tests. La fidélité est un prérequis à la validité, mais elle ne la garantit pas. Il est par conséquent important de prouver la fidélité et la validité de l'évaluation de manière indépendante.

La validité de construit mesure le degré de pertinence des inférences fondées sur l'observation ou la mesure. Pour les évaluations en milieu de travail, la validité conceptuelle est particulièrement importante car elle : a) confère une légitimité et une acceptabilité à la mesure standardisée du comportement des employés, b) facilite la compréhension des implications organisationnelles des résultats, et c) permet aux organisations et aux individus d'adapter les interventions afin de répondre aux objectifs de performance souhaités. Pour le **TEST DE RAISONNEMENT 20'**, la validité

conceptuelle a été établie en utilisant deux approches: les corrélations de facteurs et l'analyse factorielle de confirmation.

Les facteurs du **TEST DE RAISONNEMENT 20'** mesurent les capacités de raisonnement. Cependant, chaque facteur mesure un type de compétence de raisonnement et, par conséquent, la personne se voit attribuer 3 notes distinctes pour chaque type de compétence. Par conséquent, si une personne obtient un score élevé sur un facteur, cela ne signifie pas qu'elle obtiendra un score élevé sur les deux autres facteurs.

Le coefficient de corrélation est une méthode numérique permettant de quantifier la relation entre les variables. Il est compris entre -1 et 1. Plus la valeur est proche de 1, plus la relation est forte. Le coefficient de corrélation n'implique pas de causalité, cela signifie simplement qu'il y a une association entre les variables. Traditionnellement, les corrélations de facteurs sont mesurées à l'aide de la corrélation de Pearson (1895).

Tableau 8. Corrélations inter-facteurs.

	Raisonnement figural	Raisonnement numérique
Raisonnement numérique	0,56	
Raisonnement verbal	0,49	0,44

Les résultats présentés dans le tableau 8 montrent qu'il existe une relation positive entre les 3 facteurs. Entre le *Raisonnement figural* et le *Raisonnement numérique*, il existe une relation modérée ascendante (positive). Entre le *Raisonnement verbal* et le *Raisonnement figural* ainsi qu'entre le *Raisonnement verbal* et le *Raisonnement numérique*, il existe une faible relation linéaire ascendante (positive).

Il faut s'attendre à une corrélation positive dans une évaluation du raisonnement puisque les trois facteurs, *Raisonnement numérique*, *Raisonnement verbal* et *Raisonnement figural*, contribuent à la mesure d'un facteur de quotient général de second ordre ou, en d'autres termes, d'une capacité intellectuelle globale.

Analyse factorielle confirmatoire (AFC)

Pour étudier les construits théoriques, ou facteurs, pouvant être représentés par un ensemble d'items, ainsi que la qualité des items individuels, des analyses factorielles confirmatoires (AFC) sont généralement réalisées. L'AFC est utilisée pour déterminer dans quelle mesure le modèle correspond aux données. Il est donc essentiel d'évaluer la qualité des indices d'ajustement. Plusieurs tests statistiques sont utilisés pour déterminer dans quelle mesure le modèle correspond aux données en AFC et SEM. Plus les indices d'ajustement rapprochés répondent aux exigences statistiques, meilleur est le modèle. La statistique du chi carré (χ^2) est considérée comme le principal indice de la qualité de l'ajustement. Cependant, en raison de problèmes liés à la dépendance à

l'égard du χ^2 (sensibilité à la taille des échantillons et inadéquation lorsque les hypothèses de distribution sont enfreintes), des mesures de l'ajustement relatif sont fortement recommandées. Le χ^2 est une mesure de l'ajustement absolu et fournit une indication fondamentale de l'adéquation de la théorie proposée aux données et ne repose pas sur une comparaison avec un modèle de base (Jöreskog & Sörbom, 1993).

Pour la validation du **TEST DE RAISONNEMENT 20'**, l'erreur quadratique moyenne de l'approximation (RMSEA, Steiger et Lind, 1980), l'indice d'ajustement comparatif de Bentler (CFI) et l'indice Tucker-Lewis (TLI) ont été choisis comme mesures supplémentaires de l'adéquation, car ils étaient les moins susceptibles d'être affectés par la taille de l'échantillon (Fan, Thompson et Wang, 1999). En outre, l'utilisation d'indices différents est nécessaire car ils reflètent différents aspects de l'adéquation au modèle (Hooper, Coughlan et Mullen, 2008). Comme χ^2 , le RMSEA est une mesure de l'adéquation absolue qui évalue la quantité d'erreur associée à un modèle proposé. Plus le RMSEA est bas, plus le modèle est robuste. Des indices de 0,005-0,075 indiquent une bonne adéquation (Hu et Bentler, 1999). C'est parce que le RMSEA concerne les résidus du modèle et qu'il évalue dans quelle mesure le modèle est inexact. Le CFI comme le TLI sont des indices de l'adéquation relative ou incrémentale qui évaluent l'amélioration globale d'un modèle proposé par rapport à un modèle indépendant où les variables observées ne sont pas corrélées (Hooper et al. 2008, Holtzman et Vezzu, 2011). Si les valeurs supérieures du CFI et du TLI (0,9 ou plus) indiquent une bonne adéquation, cela ne correspond pas forcément à un RMSEA inférieur ou égal à 0,05 (Hu et Bentler, 1999).

Tableau 9. Qualité des indices d'ajustement et des résultats du modèle.

N	χ^2	df	χ^2/df	P	RMSEA	P<=.05	CFI	TLI
7754	996,71	186	5,36	0,00	0,024	1,00	0,974	0,971

Les résultats présentés dans le tableau 9 montrent un RMSEA de 0,024 et un CFI de 0,974, ce qui indique que le **TEST DE RAISONNEMENT 20'** a un très bon ajustement de modèle et que les facteurs existent dans la structure définie.

2.4. TEST DE RAISONNEMENT 20' et égalité des chances

Les organisations sont de plus en plus mises sous pression pour veiller à ce que les outils utilisés pour aider aux prises de décisions de recrutement, de sélection, de développement, de promotion et de mesure du rendement des employés soient justes, objectifs et précis. Cette pression découle d'un bassin de candidats de mieux en mieux informés, d'une main-d'œuvre plus sophistiquée et d'une législation du travail et de normes d'essais locales et internationales strictes. L'égalité des chances est donc un élément essentiel de toute évaluation du travail moderne, et une évaluation n'est considérée comme vraiment valable que si elle convient à une utilisation avec différents groupes. Le **TEST DE RAISONNEMENT 20'** constitue une mesure juste et précise des capacités de raisonnement. Les praticiens des ressources humaines et les psychologues peuvent être certains de son application.

Équivalence de mesure

L'équivalence de mesure (EM) désigne la cohérence de la mesure entre différents groupes aux caractères spécifiques, tels que la nationalité, l'origine ethnique et le genre. Elle vise à garantir que les personnes de différents groupes ayant le même statut sur un attribut aient le même ensemble d'observations lors d'une évaluation (Drasgow, 1987; McLeary & Cruise, 2012). L'EM aide à déterminer dans quelle mesure les répondants interprètent les items de la même manière, et à savoir si les différences identifiées entre les groupes reflètent les différences réelles dans le trait mesuré (Drasgow, 1987). L'une des méthodes permettant d'évaluer l'équivalence des mesures est l'analyse du fonctionnement différentiel des items.

Analyse différentielle du fonctionnement des items (DIF)

Un item fonctionne différemment, ou affiche DIF, si les répondants de différents groupes ont des probabilités d'endosser l'item différentes, bien qu'ils aient le même niveau sur un attribut ou un trait. Quand une probabilité différente d'endosser un item reste constante à travers différents niveaux de l'attribut ou trait, le DIF est dit "uniforme". Dans ce cas, l'item concerné donne systématiquement un avantage à un groupe à tous les niveaux sur l'attribut ou trait. Inversement, quand une probabilité différente d'endosser un item varie au niveau de l'attribut ou du trait, le DIF est qualifié de "non uniforme" (Clauser et Mazor, 1998; Hanson, 1998). Plutôt qu'un avantage constant accordé au groupe de référence à travers le continuum des capacités, un avantage pour un groupe par rapport à un autre existe pour différents niveaux du trait. Par exemple, un item peut avantager le groupe de référence à l'extrémité supérieure du continuum, tout en le désavantageant à l'extrémité inférieure.

Analyse DIF en fonction du genre

L'analyse DIF a été réalisée pour les deux sexes afin de s'assurer qu'il n'y avait pas de biais envers les hommes ou les femmes et que, par conséquent, le **TEST DE RAISONNEMENT 20'** est un test juste pour les deux groupes. Pour cette analyse, l'échantillon de normes a été divisé en fonction du genre : 3793 femmes et 3961 hommes.

Tableau 10. Coefficient et taille de l'effet pour des DIF significatifs.

Factor	Coefficient	R^2	Taille de l'effet
Raisonnement figural 1	2,45	0,000	Négligeable
Raisonnement figural 2	49,04	0,000	Négligeable
Raisonnement verbal 1	26,89	0,000	Négligeable
Raisonnement verbal 2	27,31	0.000	Négligeable

L'analyse a identifié 4 des 21 items comme marqués d'un DIF, en fonction du sexe. Le tableau 10 indique le coefficient DIF ainsi que la taille de l'effet R^2 . Selon la classification de Zumbo et Thomas (1997) de la taille de l'effet, tout R^2 inférieur à 0,13 est considéré comme négligeable. Aucun des 4 éléments marqués pour DIF n'a montré une taille d'effet de 0,13 ou plus. Par conséquent, on peut considérer que le test de raisonnement 20' ne contient aucun élément fonctionnant différemment selon le genre.

Tests T et ANOVAS

Bien que 4 éléments aient été marqués avec DIF, leur impact sur l'évaluation globale est négligeable, ce qui signifie que l'évaluation ne privilégie pas le genre. Par conséquent, toute différence significative peut être attribuée à de réelles différences entre les groupes, plutôt qu'à des biais de l'évaluation. Pour explorer davantage les différences trouvées entre les groupes, un test t pour échantillon indépendant a été réalisé entre les groupes. Les tests t et ANOVAS ont été réalisés entre les sexes et les groupes d'âge (voir tableaux 11 et 12).

Analyse en fonction du genre

Afin de vérifier les différences de scores entre les sexes, un test t pour échantillon indépendant a été réalisé pour comparer tous les facteurs du **TEST DE RAISONNEMENT 20'** entre les femmes (n = 3793) et les hommes (n = 3961) (voir le tableau 11).

Tableau 11. Test t pour les facteurs en fonction du genre.

Facteur	Femmes	Hommes	T.	df	P
Raisonnement figural	6,22	6,41	-3,70	7752	0,00
Raisonnement numérique	5,55	5,97	-6,97	7752	0,00
Raisonnement verbal	6,35	6,40	-1,14	7752	0,25

Les tests t ont révélé qu'il y avait une différence significative, pour les facteurs *Raisonnement numérique* et *Raisonnement figural*, entre les hommes et les femmes, en faveur des hommes.

Analyse en fonction du groupe d'âge

Afin de vérifier les différences de scores entre les groupes d'âge, une ANOVA a été réalisée pour comparer tous les facteurs du **TEST DE RAISONNEMENT 20'** entre cinq groupes d'âge : 18-25 (n = 2549), 26-33 (n = 2563), 34-41 (n = 1522), 42-49 (n = 596) et 50+ (n = 323) (voir tableau 12).

Tableau 12. ANOVA par groupe d'âge.

Facteur	18-25	26-33	34-41	42-49	50+	F	df	p-value
Raisonnement figural	6,66	6,28	6,03	6,16	5,85	$1,966 \cdot 10^{35}$	4,00	0,00
Raisonnement numérique	5,88	5,76	5,68	5,86	5,65	$2,387 \cdot 10^{35}$	4,00	0,00
Raisonnement verbal	6,75	6,10	6,08	6,63	6,82	$1,064 \cdot 10^{36}$	4,00	0,00

L'ANOVA a révélé des différences significatives pour tous les facteurs. Une analyse post-hoc a été réalisée à l'aide de tests t pairés (correction de Bonferroni) afin de voir où ces différences se produisaient.

Pour le *Raisonnement figural* :

Les individus de 50 ans et plus ($M = 5,85$) ont obtenu des résultats significativement inférieurs à ceux des individus de 18 à 25 ans ($M = 6,66$, $p < 0,001$) et de 26 à 33 ans ($M = 6,28$, $p = 0,028$);

Les personnes ayant entre 42 et 49 ans ($M = 6,16$) montrent des résultats significativement inférieurs à ceux des 18-25 ans ($M = 6,66$, $p < 0,001$);

Les participantes ayant entre 34 et 41 ans ($M = 6,03$) ont obtenu des résultats inférieurs à ceux des 18-25 ans ($M = 6,66$, $p < 0,001$) et des 26-33 ans ($M = 6,28$, $p = 0,017$);

Enfin, les 26-33 ans ($M = 6,28$) ont également obtenu des résultats inférieurs à ceux des 18-25 ans ($M = 6,66$, $p < 0,001$).

Pour le *Raisonnement numérique* :

Les individus âgés de 18 à 25 ans ($M = 5,88$) ont obtenu des résultats significativement supérieurs à ceux des autres groupes ($p < 0,001$).

Pour le *Raisonnement verbal* :

Les individus âgés de 50 ans et plus ($M = 6,82$) ont obtenu des résultats significativement supérieurs à ceux des personnes ayant entre 26 et 33 ans ($M = 6,10$, $p < 0,001$) et entre 34 et 41 ans ($M = 6,08$, $p < 0,001$);

Les individus ayant entre 42 et 49 ans ($M = 6,63$) montrent des résultats significativement supérieurs à ceux des participants ayant entre 26 et 33 ans ($M = 6,10$, $p < 0,001$) et entre 34 et 41 ans ($M = 6,08$, $p < 0,001$);

Les participantes âgées de 34 à 41 ans ($M = 6,08$) ont obtenu des résultats significativement inférieurs à ceux des 18-25 ans ($M = 6,75$, $p < 0,001$), tout comme les individus de 26 à 33 ans ($M = 6,10$, $p < 0,001$).

3. Informations supplémentaires et Considérations

Le **TEST DE RAISONNEMENT 20'** a été développé principalement pour guider les décisions des ressources humaines en matière de recrutement, mais son utilisation est multiple pour les paramètres organisationnels. Il est important que les normes, les scores et les interprétations appropriés soient utilisés afin de faciliter la sélection, le développement de carrière, le coaching et la gestion des performances. Une interprétation précise des tests psychométriques aide les administrateurs et les candidats à : a) placer l'évaluation dans son contexte ; b) comparer efficacement les candidats entre les capacités cibles ; c) identifier avec précision les domaines de développement ; d) optimiser l'adéquation entre les exigences du poste et le potentiel du candidat ; e) promouvoir l'organisation en tant que lieu de travail juste et objectif. Le présent chapitre fournit aux professionnels des ressources humaines des lignes directrices pour comprendre et interpréter avec succès les scores des candidats au **TEST DE RAISONNEMENT 20'** pour différents objectifs organisationnels.

Le logiciel de génération de rapports pour le **TEST DE RAISONNEMENT 20'** fournit des graphiques et des descriptions pour les trois facteurs liés au score du candidat. Le rapport ne doit être interprété que par des professionnels ayant suivi la formation agréée par *Central Test* et connaissant ce manuel. La formation et la connaissance du processus d'évaluation sont essentielles car les notes doivent être interprétées à la lumière d'informations relatives à la fiabilité, à la validité, à l'erreur type de mesure et à toute adaptation apportée à l'évaluation ou à la séance d'évaluation. Les scores doivent également être présentés dans le contexte de l'évaluation et de manière claire, significative et appropriée pour le public cible. Il est également important de noter que toutes les évaluations effectuées par *Central Test* ont été développées uniquement à des fins de ressources humaines. Par conséquent, aucune conclusion clinique ne peut être tirée des résultats, pas même pour des scores extrêmes.

Le recrutement et la sélection constituent une fonction organisationnelle importante qui peut prendre beaucoup de temps et coûter très cher. De plus, les mauvaises décisions de recrutement ont des effets supplémentaires sur la productivité, les capacités et la durabilité de l'organisation. Par conséquent, il est recommandé aux recruteurs, avant d'administrer le **TEST DE RAISONNEMENT 20'**, de disposer d'une description de poste détaillée, fondée sur une analyse des tâches demandées.

De plus, le **TEST DE RAISONNEMENT 20'** peut être utilisé conformément aux évaluations de performance, qui sont un élément clé du développement et du succès des employés. Les évaluations de performances doivent être des évaluations justes et équilibrées des réalisations et des capacités d'un employé et doivent donner l'occasion de discuter des attentes et des réalisations.

Si le **TEST DE RAISONNEMENT 20'** identifie efficacement les domaines à développer, il ne doit pas être utilisé pour suivre directement les progrès des performances d'un employé. Nous recommandons donc l'utilisation d'évaluations objectives de la performance ou d'autres évaluations en conjonction avec le **TEST DE RAISONNEMENT 20'**.

Références

- Clauser, B. E., & Mazor, K. M. (1998). Using Statistical Procedures to Identify Differentially Functioning Test Items. *Educational Measurement: Issues and Practice*, 17(1), 31-44. <http://doi.org/10.1111/j.1745-3992.1998.tb00619.x>
- Dewey, J. (1910/1933). *How we think: A restatement of the reflective thinking to the educative process*. Boston: Heath.
- Drasgow, F. (1987). Study of the measurement bias of two standardized psychological tests. *Journal of Applied Psychology*, 72(1), 19-29. <http://doi.org/10.1037/0021-9010.72.1.19>
- Fan, X., Thompson, B., & Wang, L. (1999). Effects of sample size, estimation methods, and model specification on structural equation modeling fit indexes. *Structural Equation Modeling: A Multidisciplinary Journal*, 6(1), 56-83.
- Hanson, B. A. (1998). Uniform DIF and DIF Defined by Differences in Item Response Functions. *Journal of Educational and Behavioral Statistics*, 23(3), 244-253. <http://doi.org/10.3102/10769986023003244>
- Holtzman, S., & Vezzu, S. (2011). *Confirmatory Factor Analysis and Structural Equation Modeling of Noncognitive Assessments using PROC CALIS*. <http://www.lexjansen.com/nesug/nesug11/sa/sa07.pdf>
- Hooper, D., Coughlan, J., & Mullen, M. (2008). Structural Equation Modelling: Guidelines for Determining Model Fit. *Electronic Journal of Business Research Methods*, 6(1), 53-60.
- Hu, L., & Bentler, P. M. (1999). Cut-off criteria for fit indexes in covariance structure analysis: Conventional criteria versus new alternatives. *Structural Equation Modeling: A Multidisciplinary Journal*, 6(1), 1-55. doi.org/10.1080/10705519909540118
- Jöreskog, K. G., & Sörbom, D. (1993). LISREL 8: Structural Equation Modeling with the SIMPLIS Command Language. *Scientific Software International*.
- Lila, M., Oliver, A., Catalá-Miñana, A., Galiana, L., & Gracia, E., (2014). The intimate partner violence responsibility attribution scale (IPVRAS). *The European Journal of Psychology Applied to Legal Context*, 6(1), 29-36.
- McLeary, C., & Cruise, P. (2012). Modeling Organisational Trust: Addressing Challenges of Measurement Invariance. In *Proceedings of the 11th European Conference on Research Methods* (p. 470). Academic Conferences Limited.
- Nunnally, J. C. (1978). *Psychometric theory*. New York: McGraw-Hill.

- Pearson, K. (1895). Contributions to the Mathematical Theory of Evolution. II. Skew Variation in Homogeneous Material. *Philosophical Transactions of the Royal Society of London: Mathematical, Physical and Engineering Sciences*, 186, 343-414.
<http://doi.org/10.1098/rsta.1895.0010>
- Schmidt, F. L., & Hunter, J. E. (1996). Intelligence and Job Performance: Economic and Social Implications. *Psychology Public Policy and Law*, Vol. 2, 447-472.
- Steiger, J. H., & Lind, J. C. (1980). *Statistically based tests for the number of common factors*. In annual meeting of the Psychometric Society, Iowa City, IA, 758.
- Trilling, B., & Fadel, C. (2009). *21st century skills: Learning for life in our times: Learning for life in our times*. San Francisco: John Wiley & Sons.
- Wilcox, S., Schoffman, D. E., Dowda, M., & Sharpe, P. A. (2014). Psychometric properties of the 8-item english arthritis self-efficacy scale in a diverse sample. *Arthritis*.
- Zumbo, B. D., & Thomas, D. R. (1997). *A measure of effect size for a model-based approach for studying DIF*. Prince George, Canada: Edgeworth Laboratory for Quantitative Behavioral Science, University of Northern British Columbia.

NOUS CONTACTER

Service Client: **service.client@centraltest.com**

Formation et Consulting: **consulting@centraltest.com**



www.centraltest.com

Central Test – Empowering Talent © 2020

Tous droits réservés.